

WHAT IS CLAIMED IS:

1. A speech recognition method, including:
 - creating speech data on which different types of noise have been superposed respectively;
 - 5 eliminating the noise by a predetermined noise elimination method from each of the speech data on which the noise has been superposed;
 - creating acoustic models corresponding to each of the noise types;
 - storing the acoustical models using feature vectors of each of the speech data which have undergone the noise elimination;
 - 10 when a speech recognition is performed, determining the type of a noise superposed on speech data to be recognized;
 - selecting a corresponding acoustic model from the acoustic models corresponding to each of the noise types based on the result of the determination;
 - eliminating the noise by the predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and
 - 15 performing a speech recognition on the feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.
2. The speech recognition method according to Claim 1, wherein the noise elimination method is at least one of a spectral subtraction method and a
 - 20 continuous spectral subtraction method, and the acoustic models are created by eliminating the noise by the at least one of the spectral subtraction method and the continuous spectral subtraction method from each of the speech data on which the different types of noise have been superposed, obtaining the feature vectors of each of the speech data which have undergone the noise elimination, and using the feature
 - 25 vectors;
 - when a speech recognition is performed, a first speech feature analysis is performed to obtain frequency-domain feature data of the speech data on which the noise has been superposed;
 - a determination is made whether the speech data is a noise segment or
 - 30 a speech segment based on the result of the feature analysis, and when a noise segment is detected, the feature data thereof is stored, whereas when a speech segment is detected, the type of the noise superposed is determined based on the feature data having been stored and a corresponding acoustic model is selected from

the acoustic models corresponding to each of the noise types based on the result of the determination;

the noise is eliminated by the at least one of the spectral subtraction method and the continuous spectral subtraction method from the speech data to be recognized on which the noise has been superposed; and

a second feature analysis is performed on the speech data which has undergone the noise elimination to obtain feature data required in the speech recognition, and a speech recognition is performed on the result of the feature analysis based on the selected acoustic model.

3. The speech recognition method according to Claim 1, wherein the noise elimination method is a cepstrum mean normalization method, and the acoustic models are created by eliminating the noise by the cepstrum mean normalization method from each of the speech data on which the different types of noise have been superposed and using the feature vectors of the speech data obtained thereby;

when a speech recognition is performed, a first speech feature analysis is performed on the speech data to be recognized on which the noise has been superposed to obtain a feature vector representing cepstrum coefficients;

a determination is made whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and when a noise segment is detected, the feature vector thereof is stored, and when a speech segment is detected, the feature data of the speech segment from a beginning through an end thereof is stored, the type of the noise superposed is determined based on the feature vector of the noise segment having been stored, and an acoustic model is selected from the acoustic models corresponding to each of the noise types based on the result of the determination;

the noise is eliminated by the cepstrum mean normalization method from the speech segment on which the noise has been superposed using the feature vector of the speech segment having been stored; and

a speech recognition is performed on the feature vector after the noise elimination based on the selected acoustic model.

4. The speech recognition method according to Claim 1, wherein the acoustic models correspond to types of noise and additionally a plurality of S/N ratios for each of the noise types, and the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which

noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of the speech data by a predetermined noise elimination method, and using the feature vectors of each of the speech data which have undergone the noise elimination.

5 5. The speech recognition method according to Claim 4, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech segment, and an
10 acoustic model is selected based on a noise type determined and the S/N ratio obtained.

 6. A speech recognition method, wherein speech data on which different types of noise have been superposed respectively are created, the noise is eliminated by a spectral subtraction method or a continuous spectral subtraction method from
15 each of the speech data on which the different types of noise have been superposed, a cepstrum mean normalization method is applied to each of the speech data which have undergone the noise elimination to obtain feature vectors of a speech segment, and acoustic models corresponding to each of the noise types are created and stored based on the feature vectors;

20 when a speech recognition is performed, a first speech feature analysis is performed to obtain frequency-domain feature data of speech data to be recognized; a determination is made whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and when a noise segment is detected, the feature vector thereof is stored, and when a speech segment is
25 detected, the noise is eliminated from the speech segment by the spectral subtraction method or the continuous spectral subtraction method;

 a second speech feature analysis is performed on the speech segment data which has undergone the noise elimination to obtain cepstrum coefficients, and the feature vector of the speech segment is stored;

30 when the speech segment has terminated, the type of the noise superposed is determined based on the feature data of the noise segment having been stored, and an acoustic model is selected from the acoustic models corresponding to each of the noise types;

the cepstrum mean normalization method is applied to the feature vector of the speech segment on which the noise has been superposed, using the feature vector of the speech segment having been stored, to obtain the feature vector of the speech segment; and

5 a speech recognition is performed on the feature vector obtained by the cepstrum mean normalization method based on the selected acoustic model.

7. The speech recognition method according to Claim 6, wherein the acoustic models correspond to types of noise in addition to a plurality of S/N ratios for each of the noise types, and the acoustic models corresponding to the plurality of
10 S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of the speech data by the spectral subtraction method or the continuous spectral subtraction method, and using the feature vectors of each of the speech data obtained by applying the cepstrum mean
15 normalization method to each of the speech data which have undergone the noise elimination.

8. The speech recognition method according to Claim 7, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the
20 speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech segment, and an acoustic model is selected based on the noise type determined and the S/N ratio obtained.

9. A speech recognition method, comprising:
25 creating speech data on which a particular type of noise with different S/N ratios have been superposed respectively;
eliminating the noise by a predetermined noise elimination method from each of the speech data;
creating and storing acoustic models corresponding to each of the S/N
30 ratios using feature vectors of each of the speech data which have undergone the noise elimination;
when a speech recognition is performed, determining the S/N ratio of a noise superposed on speech data to be recognized;

selecting a corresponding acoustic model from the acoustic models corresponding to each of the S/N ratios based on the result of the determination;

eliminating the noise by the predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and

performing a speech recognition on the feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.

10. The speech recognition method according to Claim 9, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method.

11. The speech recognition method according to Claim 9, wherein the noise elimination method is a cepstrum mean normalization method.

12. A storage medium storing a speech recognition program, the speech recognition program comprising the steps of:

creating speech data on which different types of noise have been superposed respectively, eliminating the noise by a predetermined noise elimination method from each of the speech data on which the noise has been superposed, and creating acoustic models corresponding to each of the noise types using the feature vectors obtained by analyzing the features of each of the speech data which have undergone the noise elimination, and storing the acoustic models in acoustic model storage device;

determining the type of a noise superposed on speech data to be recognized, and selecting a corresponding acoustic model from the acoustic models stored in said acoustic model storage device based on the result of the determination;

eliminating the noise by the predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and performing a speech recognition on the feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.

13. The storage medium storing a speech recognition program according to Claim 12, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method, and the acoustic models are created by eliminating the noise by the spectral subtraction method or the continuous spectral subtraction method from each of the speech data on which the different types of noise have been superposed, obtaining the feature vectors of each of

the speech data which have undergone the noise elimination, and using the feature vectors, the process of speech recognition comprises the steps of:

performing a first speech feature analysis to obtain frequency-domain feature data of the speech data on which the noise has been superposed;

5 determining whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, storing the feature data thereof when a noise segment is detected, whereas when a speech segment is detected, determining the type of the noise superposed based on the feature data having been stored and selecting a corresponding acoustic model from the acoustic models
10 corresponding to each of the noise types based on the result of the determination;

eliminating the noise by the spectral subtraction method or the continuous spectral subtraction method from the speech data to be recognized on which the noise has been superposed; and

performing a second feature analysis on the speech data which has
15 undergone the noise elimination to obtain feature data required in the speech recognition, and performing a speech recognition on the result of the feature analysis based on the selected acoustic model.

14. The storage medium storing a speech recognition program according to Claim 12, wherein the noise elimination method is a cepstrum mean normalization
20 method, and the acoustic models are created by eliminating the noise by the cepstrum mean normalization method from each of the speech data on which the different types of noise have been superposed and using the feature vectors of the speech data obtained thereby, the process of speech recognition comprises the steps of:

performing a first speech feature analysis on the speech data to be
25 recognized on which the noise has been superposed to obtain a feature vector representing cepstrum coefficients;

determining whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and storing the feature vector thereof when a noise segment is detected, whereas when a speech segment is detected,
30 storing the feature data of the speech segment from the beginning through the end thereof, determining the type of the noise superposed based on the feature vector of the noise segment having been stored, and selecting an acoustic model from the acoustic models corresponding to each of the noise types based on the result of the determination;

eliminating the noise by the cepstrum mean normalization method from the speech segment on which the noise has been superposed using the feature vector of the speech segment having been stored, and

performing a speech recognition on the feature vector after the noise
5 elimination based on the selected acoustic model.

15. The storage medium storing a speech recognition program according to Claim 12, wherein the acoustic models correspond to types of noise in addition to a plurality of S/N ratios for each of the noise types, and the acoustic models
10 corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of the speech data by a predetermined noise elimination method, and using the feature
vectors of each of the speech data which have undergone the noise elimination.

16. The storage medium storing a speech recognition program according to
15 Claim 15, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech segment, and an acoustic model is selected based on the noise type determined and
20 the S/N ratio obtained.

17. A storage medium storing a speech recognition program, the speech
recognition program comprising the steps of:

creating speech data on which different types of noise have been
superposed respectively, eliminating the noise by at least one of a spectral subtraction
25 method and a continuous spectral subtraction method from each of the speech data on which the different types of noise have been superposed, applying a cepstrum mean normalization method to each of the speech data which have undergone the noise elimination to obtain the feature vectors of a speech segment, and creating acoustic
models corresponding to each of the noise types based on the feature vectors and
30 storing the acoustic models in acoustic model storage means;

performing a first speech feature analysis to obtain frequency-domain
feature data of speech data to be recognized on which a noise has been superposed;

determining whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and storing the feature vector thereof when a noise segment is detected;

5 eliminating the noise from the speech segment by the at least one of the spectral subtraction method and the continuous spectral subtraction method when a speech segment is detected;

performing a second speech feature analysis on the speech segment data which has undergone the noise elimination to obtain cepstrum coefficients, and storing the feature vector of the speech segment;

10 when the speech segment has terminated, determining the type of the noise superposed based on the feature data of the noise segment having been stored, and selecting an acoustic model from the acoustic models corresponding to each of the noise types;

15 applying the cepstrum mean normalization method to the feature vector of the speech segment on which the noise has been superposed, using the feature vector of the speech segment having been stored, to obtain the feature vector of the speech segment; and

performing a speech recognition on the feature vector obtained by the cepstrum mean normalization method based on the selected acoustic model.

20 18. The storage medium storing a speech recognition program according to Claim 17, wherein the acoustic models correspond to types of noise in addition to a plurality of S/N ratios for each of the noise types, and the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of
25 the speech data by the at least one of the spectral subtraction method and the continuous spectral subtraction method, and using the feature vectors of each of the speech data obtained by applying the cepstrum mean normalization method to each of the speech data which have undergone the noise elimination.

30 19. The storage medium storing a speech recognition program according to Claim 17, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech

segment, and an acoustic model is selected based on the noise type determined and the S/N ratio obtained.

20. A storage medium storing a speech recognition program, the speech recognition program comprising the steps of:

5 creating speech data on which a particular type of noise with different S/N ratios have been superposed respectively, eliminating the noise by a predetermined noise elimination method from each of the speech data, and creating acoustic models corresponding to each of the S/N ratios using the feature vectors of each of the speech data which have undergone the noise elimination and storing the acoustic models in acoustic model storage device;

 determining the S/N ratio of a noise superposed on speech data to be recognized, and selecting a corresponding acoustic model from the acoustic models corresponding to each of the S/N ratios based on the result of the determination;

 eliminating the noise by the predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and
15 performing a speech recognition on the feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.

21. The storage medium storing a speech recognition program according to Claim 20, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method.

22. The storage medium storing a speech recognition program according to Claim 20, wherein the noise elimination method is a cepstrum mean normalization method.

23. A speech recognition apparatus, comprising:
25 acoustic models corresponding to each of different types of noise, created by generating speech data on which the different types of noise have been superposed respectively, eliminating the noise by a predetermined noise elimination method from each of the speech data on which the different types of noise have been superposed, and using the feature vectors of each of the speech data which have
30 undergone the noise elimination;

 an acoustic model storage device that stores the acoustic models;

 a noise determination device that determines the type of a noise superposed on speech data to be recognized;

an acoustic model selection device that selects a corresponding acoustic model from the acoustic models corresponding to each of the noise types based on the result of the determination;

a noise elimination device that eliminates the noise by the
5 predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and

a speech recognition device that performs a speech recognition on the feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.

10 24. The speech recognition apparatus according to Claim 23, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method, and the acoustic models are created by eliminating the noise by the at least one of the spectral subtraction method and the continuous spectral subtraction method from each of the speech data on which the
15 different types of noise have been superposed, obtaining the feature vectors of each of the speech data which have undergone the noise elimination, and using the feature vectors, the speech recognition apparatus comprises:

an acoustic model storage device that stores the acoustic models thus created;

20 a first speech feature analysis device that performs a first speech feature analysis to obtain frequency-domain feature data of the speech data on which the noise has been superposed;

a noise segment/speech segment determination device that determines whether the speech data is a noise segment or a speech segment based on the result of
25 the feature analysis, and when a noise segment is detected, storing the feature data thereof in feature data storage means;

a noise type determination device that determines the type of noise the type of the noise superposed based on the feature data having been stored when a speech segment is detected;

30 an acoustic model selection device that selects a corresponding acoustic model from the acoustic models corresponding to each of the noise types based on the result of the determination;

a noise elimination device that eliminates the noise by the at least one of the spectral subtraction method and the continuous spectral subtraction method from the speech data to be recognized on which the noise has been superposed;

a second speech feature analysis device that performs a second feature analysis on the speech data which has undergone the noise elimination to obtain feature data required in the speech recognition; and

a speech recognition device that performs a speech recognition on the result of the feature analysis based on the selected acoustic model.

25. The speech recognition apparatus according to Claim 23, wherein the noise elimination method is a cepstrum mean normalization method, and the acoustic models are created by eliminating the noise by the cepstrum mean normalization method from each of the speech data on which the different types of noise have been superposed and using the feature vectors of the speech data obtained thereby, the speech recognition apparatus comprises:

an acoustic model storage device that stores the acoustic models thus created;

a feature analysis device that performs a first speech feature analysis on the speech data to be recognized on which the noise has been superposed to obtain a feature vector representing cepstrum coefficients;

a noise segment/speech segment determination device that determines whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and storing the feature vector thereof in feature data storage means when a noise segment is detected, whereas when a speech segment is detected, storing the feature data of the speech segment from a beginning through a end thereof in the feature data storage device;

a noise type determination device that determines the type of the noise superposed based on the feature vector of the noise segment having been stored in the feature data storage device;

an acoustic model selection device that selects a corresponding acoustic model from the acoustic models corresponding to each of the noise types based on the result of the determination;

a noise elimination device that eliminates the noise by the cepstrum mean normalization method from the speech segment on which the noise has been superposed using the feature vector of the speech segment having been stored; and

a speech recognition device that performs a speech recognition on the feature vector after the noise elimination based on the selected acoustic model.

26. The speech recognition apparatus according to Claim 23, wherein the acoustic models correspond to types of noise in addition to a plurality of S/N ratios for each of the noise types, and the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of the speech data by a predetermined noise elimination method, and using the feature vectors of each of the speech data which have undergone the noise elimination.

27. The speech recognition apparatus according to Claim 26, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the noise type determination device obtains the S/N ratio from the magnitude of the noise in the noise segment and the magnitude of the speech in the speech segment, and the acoustic model selection device selects an acoustic model based on the noise type determined and the S/N ratio obtained.

28. A speech recognition apparatus, comprising:
acoustic models corresponding to each of different types of noise, created by generating speech data on which the different types of noise have been superposed respectively, eliminating the noise by a spectral subtraction method or a continuous spectral subtraction method from each of the speech data on which the different types of noise have been superposed, applying a cepstrum mean normalization method to each of the speech data which have undergone the noise elimination to obtain the feature vectors of a speech segment, and using the feature vectors;

an acoustic model storage device that stores the acoustic models;
a first speech feature analysis device that performs a first speech feature analysis to obtain frequency-domain feature data of speech data to be recognized;

a noise segment/speech segment determination device that determines whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and stores the feature vector thereof in feature data storage means when a noise segment is detected;

a noise elimination device that eliminates the noise from the speech segment by the spectral subtraction method or the continuous spectral subtraction method when a speech segment is detected;

5 a second speech feature analysis device that performs a second speech feature analysis on the speech segment data which has undergone the noise elimination to obtain cepstrum coefficients, and storing the feature vector of the speech segment in the feature data storage device;

10 a noise type determination device that determines, when the speech segment has terminated, the type of the noise superposed based on the feature data of the noise segment having been stored;

an acoustic model selection device that selects a corresponding acoustic model from the acoustic models corresponding to each of the noise types;

15 a cepstrum mean normalization operation device that applies the cepstrum mean normalization method to the feature vector of the speech segment on which the noise has been superposed, using the feature vector of the speech segment having been stored, to output the feature vector of the speech segment; and

a speech recognition device that performs a speech recognition on the feature vector based on the selected acoustic model.

20 29. The speech recognition apparatus according to Claim 28, wherein the acoustic models correspond to types of noise in addition to a plurality of S/N ratios for each of the noise types, and the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed, eliminating the noises from each of the speech data by the
25 spectral subtraction method or the continuous spectral subtraction method, and using the feature vectors of each of the speech data obtained by applying the cepstrum mean normalization method to each of the speech data which have undergone the noise elimination.

30 30. The speech recognition apparatus according to Claim 29, wherein when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the noise type determination device obtains the S/N ratio from a magnitude of the noise in a noise segment and a magnitude of a speech in

the speech segment, and the acoustic model device means selects an acoustic model based on the noise type determined and the S/N ratio obtained.

31. A speech recognition apparatus, comprising:

- 5 acoustic models corresponding to each of different S/N ratios for a particular type of noise, created by generating speech data on which the particular type of noise with the different S/N ratios have been superposed respectively, eliminating the noise by a predetermined noise elimination method from each of the speech data, and using the feature vectors of each of the speech data which have undergone the noise elimination;
- 10 an acoustic models storage device that stores the acoustic models;
a S/N ratio determination means for determining the S/N ratio of a noise superposed on speech data to be recognized;
an acoustic model selection device that selects a corresponding acoustic model from the acoustic models corresponding to each of the S/N ratios
15 based on the result of the determination;
a noise elimination device that eliminates the noise by the predetermined noise elimination method from the speech data to be recognized on which the noise has been superposed; and
a speech recognition device that performs a speech recognition on the
20 feature vector of the speech data which has undergone the noise elimination based on the selected acoustic model.

32. The speech recognition apparatus according to Claim 31, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method.

25 33. The speech recognition apparatus according to Claim 31, wherein the noise elimination method is a cepstrum mean normalization method.